УДК 519.766.23

О СВОЙСТВАХ ЯЗЫКОВ, ЗАДАВАЕМЫХ МУЛЬТИМОДАЛЬНЫМИ КАТЕГОРИАЛЬНЫМИ ГРАММАТИКАМИ ЗАВИСИМОСТЕЙ

Карлов Б.Н.
Кафедра информатики

Поступила в редакцию 05.09.2011, после переработки 12.09.2011.

В работе продолжается изучение категориальных грамматик зависимостей (КГЗ). Рассматривается мультимодальный вариант этих грамматик с запретами на пересечение дальних связей (ммКГЗ). Доказывается, что множество ммКГЗ-языков замкнуто относительно операции итерации и является абстрактным семейством языков. Устанавливается, что класс ммКГЗ-языков замкнут относительно пересечения, но не замкнут относительно проекции. Построен пример неполулинейного ммКГЗ-языка. Установлено, что существует ммКГЗ с NP-полной проблемой принадлежности.

In the paper we continue to investigate the categorial dependency grammars (CDG). A multimodal version of these grammars with negative mode pairing rules (mmCDG) is considered. It is proved that the set of mmCDG-languages is closed under iteration and it is an abstract family of languages. We also establish that the class of mmCDG-languages is closed under intersection, but it is not closed under projection. An example of a non-semilinear mmCDG-language is presented. It is established that there exists a mmCDG with NP-complete membership problem.

Ключевые слова: формальные грамматики, категориальные грамматики, структура зависимостей, абстрактные семейства языков, неполулинейные множества, сложность проблемы принадлежности.
Keywords: formal grammars, categorial grammars, dependency structure, abstract families of languages, non-semilinear sets, complexity of membership problem.

1. Введение

Теории синтаксиса естественных языков, основанные на понятии *зависимости*, имеют давнюю традицию, восходящую к средним векам. Теньер [12] впервые систематически описал структуру предложения в терминах именованных отношений между словами (*зависимостей*). Когда два слова w_1 и w_2 связаны в предложении посредством зависимости d (обозначение $w_1 \xrightarrow{d} w_2$), w_1 является *главным словом*, а w_2 — *зависимым словом.* Содержательно, зависимость d задаёт ограничения на грамматические и лексические свойства w_1 и w_2, на их порядок, контекст и т.п., которые вместе означают, что “w_1 управляет w_2”. Например, в структуре зависимостей предложения *“Летом здесь играют дети”*, приведённой на рис. 1, отношение *играют* $\xrightarrow{\textit{пред}}$ *дети* показывает предикатную зависимость между сказуемым *играют* и подлежащим *дети.*

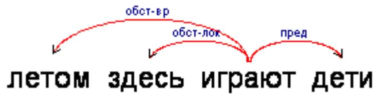

Рис. 1: Пример проективной структуры зависимостей

В этом предложении, как и в большинстве обычных предложений русского языка, структура зависимостей *проективная*, в частности, зависимости в структуре не пересекаются. Большинство грамматик, порождающих деревья зависимостей, имеют дело только с проективными структурами. С другой стороны, достаточно часто встречаются предложения, имеющие *непроективные* структуры зависимостей.

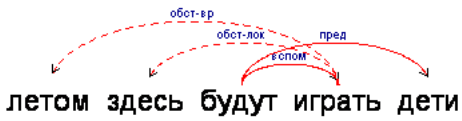

Рис. 2: Пример непроективной структуры зависимостей

Например, использование будущего времени в предложении *“Летом здесь будут играть дети”* приводит к появлению двух разрывных зависимостей *играть* $\xrightarrow{\textit{обст-вр}}$ *летом* и *играть* $\xrightarrow{\textit{обст-лок}}$ *здесь*, показанных на рис. 2.

Отметим, что кс-грамматики (см. [1, 3]), как и обычные категориальные грамматики (см. [3, 6]) неспособны обнаруживать в подобных предложениях разрывные составляющие. Для синтаксического анализа предложений с разрывными зависимостями М.И. Дехтярь и А.Я. Диковский в [7, 8] определили новые виды грамматик — *категориальные грамматики зависимостей (КГЗ)* и *обощенные категориальные грамматики зависимостей (оКГЗ).* Эти грамматики работают как грамматики-распознаватели и структуру предложения раскрывают подобно грамматикам зависимостей. Дальние (разрывные) связи обрабатываются в них с помощью так называемых поляризованных валентностей: положительных, отвечающих за слово-хозяин, и отрицательных, отвечающих за подчиненное слово. В указанных работах получены результаты о выразительной силе КГЗ и оКГЗ, предложены алгоритмы их анализа. В работе [5] автором были построены нормальные формы для оКГЗ и определены счётчиковые МП-автоматы, распознающие оКГЗ-языки.

В этой статье мы продолжаем изучение категориальных грамматик зависимостей. В ней рассматривается мультимодальный вариант этих грамматик ммКГЗ ([9]), в которых для каждого типа поляризованных валентностей можно определить свой способ сокращения. В частности, мы исследуем ммКГЗ с запретами на пересечение дальних связей. Такие запреты встречаются в естественных языках.

Например, в предложениях с прямой речью связи слов внутри кавычек не дожны пересекать связь между кавычками. С формальной точки зрения такое расширение оКГЗ увеличивает класс анализируемых языков и позволяет установить для него новые свойства замкнутости.

В разделе 2 мы приводим точные определения ммКГЗ с запретами на пересечение дальних связей и класса задаваемых этими грамматиками языков. В разделе 3 рассматривается вопрос о замкнутости класса ммКГЗ-языков относительно итерации Клини. Наличие в оКГЗ итеративных категорий побудило авторов работы [8] к утверждению о замкнутости класса оКГЗ-языков относительно итерации. Однако непосредственное использование таких категорий не позволяет получить нужный результат и проблема замкнутости класса оКГЗ-языков относительно итерации остаётся открытой. Мы показываем, что класс языков, распознаваемых ммКГЗ с запретами, замкнут относительно итерации и пересечения и, следовательно, образует абстрактное семейство языков. В разделе 4 показано, что этот класс языков содержит неполулинейные языки. В разделе 5 рассматривается сложность распознавания языков из этого класса. Ранее в работе [7] было показано, что проблема проверки по КГЗ G и слову w, принадлежит ли w языку $L(G)$, является NP-полной. При этом при ограничении числа типов валентностей предложен полиномиальный алгоритм проверки принадлежности. Мы показываем, что в классе ммКГЗ существуют грамматики, для которых проблема принадлежности является NP-полной. При этом для NP-полноты достаточно иметь два типа валентностей.

2. Основные определения

Для формализации лингвистического понятия синтаксического типа мы будем использовать понятие *категории*. Пусть $\mathbf{C}$ — непустое множество элементарных грамматических категорий (например, сказуемое, определение, дополнение и т. п.). Элементарные категории могут быть итерированы: для $C \in \mathbf{C}$ пусть C^* означает соответствующую итеративную категорию. Множество всех итеративных категорий будем обозначать $\mathbf{C}^*$.

Для описания дальних связей в [7] вводятся понятия полярности и поляризованной валентности. *Полярностью* v называется элемент множества $V = \{\searrow, \swarrow, \nwarrow, \nearrow\}$. *Поляризованная валентность* β — это выражение вида vC, где $v \in V$, $C \in \mathbf{C}$. Пары двойственных валентностей $(\nearrow C, \searrow C)$ и $(\swarrow C, \nwarrow C)$ назовём *правильными* (они являются соответствующими “скобками”). Валентности вида $\nearrow C$ и $\swarrow C$ называются левыми, а валентности вида $\searrow C$ и $\nwarrow C$ — правыми (они являются левыми и правыми скобками соответственно). Последовательность поляризованных валентностей называется *потенциалом*. Множество всех возможных потенциалов обозначается $Pot(C)$. На множестве всех элементарных категорий вводится *функция запретов* — отображение $\pi \colon \mathbf{C} \to 2^{\mathbf{C}}$. Эта функция описывает ограничения на расположение правильных пар валентностей. Если $Y \in \pi(X)$, то пара валентностей типа X не может создать зависимость, если между ними есть валентности типа Y. Например, если $\pi_1(X) = \{Y\}$, $\pi_1(Y) = \{X\}$, то в потенциале $\theta_1 = \nearrow X \nearrow Y \searrow X \searrow Y$ нет правильных пар. Действительно, $\nearrow Y$ не позволяет сформировать пару $\nearrow X \searrow X$, а $\searrow X$ не позволяет сформировать пару $\nearrow Y \searrow Y$. Если же $\pi_2(X) = \{Y\}$, $\pi_2(Y) = \varnothing$, то то пара $\nearrow Y \searrow Y$ является правильной. Если удалить её из θ_1, то останется $\nearrow X \searrow X$ — правильная пара.

Потенциал называется *сбалансированным*, если его проекция на каждую пару двойственных поляризованных валентностей является правильным скобочным словом и существует такой порядок правильных пар, что их можно удалять из потенциала, не нарушая ограничений функции запретов. Например, потенциал $\nearrow A \nearrow B \searrow A \searrow B \nearrow A \searrow A$ сбалансирован (если $\pi(A) = \pi(B) = \varnothing$). Потенциал θ_1 также сбалансирован при функции запретов π_2. Потенциал $\nearrow A \searrow B \searrow A$ не сбалансирован независимо от функции запретов. Потенциал θ_1 с функцией π_1 также не сбалансирован.

Определение 1. *Множество локальных категорий $LCat(\mathbf{C})$ — это минимальное множество такое, что:*
1) $\mathbf{C} \cup \{\varepsilon\} \subseteq LCat(\mathbf{C})$, где ε — пустой символ;
2) если $\alpha \in LCat(\mathbf{C}), A \in \mathbf{C} \cup \mathbf{C}^*$, то $[A\backslash\alpha], [\alpha/A] \in LCat(\mathbf{C})$.

Из локальных категорий и потенциалов строятся категории.

Определение 2. *Категорией γ называется выражение вида α^θ, где $\alpha \in LCat(\mathbf{C})$, $\theta \in Pot(\mathbf{C})$. Множество всех категорий обозначим $Cat(\mathbf{C})$.*

Полагаем, что конструкторы $\backslash$ и $/$ассоциативны. Поэтому любая категория γ может быть представлена в виде $\gamma = [L_k\backslash \ldots \backslash L_1\backslash C/R_1/\ldots/R_m]^\theta$. Теперь дадим определение мультимодальной категориальной грамматики зависимостей.

Определение 3. *Мультимодальной категориальной грамматикой зависимостей (ммКГЗ) называется система $G =< W, \mathbf{C}, S, \delta, \pi >$, где:*
W — конечное множество символов,
$\mathbf{C}$ — конечное множество элементарных категорий,
S — выделенная в $\mathbf{C}$ главная категория,
δ — словарь, функция на W такая, что $\delta(w) \subseteq Cat(\mathbf{C})$ для каждого слова $w \in W$ (мы будем задавать его в виде $w \mapsto \delta(w)$),
$\pi\colon \mathbf{C} \to 2^{\mathbf{C}}$ — функция запретов.

Отметим, что часто в лингвистических работах элементы множества W называют словами, а последовательности слов — предложениями.

В ммКГЗ G после приписывания категорий символам слова происходит вывод главной категории S с помощью приведенных ниже правил сокращения. Эти правила имеют вид $\Gamma \vdash \Gamma'$, где Γ — сокращаемая строка категорий, а Γ' — строка категорий, получающаяся в результате сокращения. В приведённых ниже правилах Γ_1 и Γ_2 — это произвольные строки категорий из $Cat(\mathbf{C})^*$.

Определение 4. Правила сокращения

Правила локальной зависимости:
$L^l : \Gamma_1[C]^{\theta_1}[C\backslash\alpha]^{\theta_2}\Gamma_2 \vdash \Gamma_1[\alpha]^{\theta_1\theta_2}\Gamma_2$
$L^r : \Gamma_1[\alpha/C]^{\theta_1}[C]^{\theta_2}\Gamma_2 \vdash \Gamma_1[\alpha]^{\theta_1\theta_2}\Gamma_2,$
где $C \in \mathbf{C} \cup \{\varepsilon\}$

Правила итеративной зависимости:
$I^l : \Gamma_1[C]^{\theta_1}[C^*\backslash\alpha]^{\theta_2}\Gamma_2 \vdash \Gamma_1[C^*\backslash\alpha]^{\theta_1\theta_2}\Gamma_2$
$I^l_0 : \Gamma_1[C^*\backslash\alpha]^{\theta}\Gamma_2 \vdash \Gamma_1[\alpha]^{\theta}\Gamma_2$
$I^r : \Gamma_1[\alpha/C^*]^{\theta_1}[C]^{\theta_2}\Gamma_2 \vdash \Gamma_1[\alpha/C^*]^{\theta_1\theta_2}\Gamma_2$
$I^r_0 : \Gamma_1[\alpha/C^*]^{\theta}\Gamma_2 \vdash \Gamma_1[\alpha]^{\theta}\Gamma_2,$
где $C \in \mathbf{C} \cup \{\varepsilon\}$

Правило дальней зависимости:
$D : \Gamma_1\alpha^{\theta_1\beta\theta_2\breve{\beta}\theta_3}\Gamma_2 \vdash \Gamma_1\alpha^{\theta_1\theta_2\theta_3}\Gamma_2$,
где валентности $(\beta, \breve{\beta})$ *образуют правильную пару и* θ_2 *не содержит* β, $\breve{\beta}$, *а также* $\nearrow B$, $\nwarrow B$, $\swarrow B$, $\searrow B$ *для всех* $B \in \pi(C)$, *где* C *— тип валентности* β

При выполнении сокращения по каждому из правил в структуру зависимостей добавляется стрелка, идущая в направлении зависимой категории и имеющая такое же название, что и сократившаяся категория. L^l и L^r — это классические правила сокращения. При сокращении аргумента C они создают проективную зависимость C и выполняют конкатенацию потенциалов. I^l и I^r создают произвольное количество зависимостей C. I^l_0 и I^r_0 служат, чтобы удалить итеративный подтип C^*. Правило D сокращает две двойственные поляризованные валентности и создаёт непроективную зависимость C при условии выполнения запретов, налагаемых функцией π. Как обычно, через $\vdash^*$ обозначим рефлексивное и транзитивное замыкание отношения $\vdash$.

Рассмотрим, например, предложение *“Летом здесь будут играть дети”*. Пусть его словам приписаны следующие категории: $\delta(\text{летом}) = [\varepsilon]^{\swarrow\text{обст-вр}}$, $\delta(\text{здесь}) = [\varepsilon]^{\swarrow\text{обст-лок}}$, $\delta(\text{будут}) = [S/\text{пред}/\text{вспом}]$, $\delta(\text{играть}) = [\text{вспом}]^{\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}}$, $\delta(\text{дети}) = [\text{пред}]$. Покажем, как происходит вывод для строки категорий, приписанной всему предложению.

$[\varepsilon]^{\swarrow\text{обст-вр}}\ [\varepsilon]^{\swarrow\text{обст-лок}}\ [S/\text{пред}/\text{вспом}]\ [\text{вспом}]^{\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}}\ [\text{пред}] \vdash^{L_r}$
$[\varepsilon]^{\swarrow\text{обст-вр}}\ [\varepsilon]^{\swarrow\text{обст-лок}}\ [S/\text{пред}]^{\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}}\ [\text{пред}] \vdash^{L_r}$
$[\varepsilon]^{\swarrow\text{обст-вр}}\ [\varepsilon]^{\swarrow\text{обст-лок}}\ [S]^{\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}} \vdash^{L_l}$
$[\varepsilon]^{\swarrow\text{обст-вр}}\ [S]^{\swarrow\text{обст-лок}\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}} \vdash^{L_l}$
$[S]^{\swarrow\text{обст-вр}\swarrow\text{обст-лок}\nwarrow\text{обст-вр}\nwarrow\text{обст-лок}} \vdash^{D}$
$[S]^{\swarrow\text{обст-вр}\nwarrow\text{обст-вр}} \vdash^{D} [S]$

Результирующая структура зависимостей показана на рис. 2 (локальные зависимости показаны сплошными линиями, а дальние — пунктирными). Если бы $\pi(\text{обст-лок}) = \{\,\text{обст-вр}\,\}$, $\pi(\text{обст-вр}) = \{\,\text{обст-лок}\,\}$, то приведённый выше вывод был бы невозможен из-за нарушения запретов.

Определение 5. *Слово* $s = w_1 \ldots w_n \in W^*$ *принадлежит языку* $L(G)$, *если существует строка категорий* Γ *такая, что:*
1) $\Gamma = \gamma_1 \ldots \gamma_n$, *где* $\gamma_i \in \delta(w_i)$ *для* $i = 1, \ldots, n$,
2) $\Gamma \vdash^* [S]$.

Класс всех языков, порождаемых мультимодальными категориальными грамматиками зависимостей, обозначим $\mathcal{L}(mmCDG)$.

Заметим, что в определении не накладываются никакие ограничения на функцию запретов. Если $\pi(X) = \{\,Y\,\}$, $\pi(Y) = \varnothing$, то такая функция не создаёт никаких ограничений на допустимые структуры зависимостей, изменения касаются лишь порядка сокращений категорий. Можно рассматривать функции запретов специального вида. Функция запретов π называется *симметричной*, если для любых $X, Y \in \mathbf{C}$ $X \in \pi(Y)$ тогда и только тогда, когда $Y \in \pi(X)$. В таком случае ограничения означают, что зависимости X и Y не могут пересекаться.

Несложно видеть, что локальные правила сокращения $L^l, L^r, I^l, I^l_0, I^r, I^r_0$ действуют только на локальную часть категории, тогда как правило сокращения потенциала D затрагивает только потенциал. Это даёт возможность разделить анализ слова в ммКГЗ на два независимых теста: первый — в терминах локальных категорий и локальных правил сокращения, а второй — в терминах сбалансированности потенциала. Имеет место теорема об анализе для ммКГЗ.

Теорема (об анализе) *Слово $s = w_1 w_2 \dots w_n$ принадлежит языку $L(G)$, задаваемому ммКГЗ $G =< W, \mathbf{C}, S, \delta, \pi >$, тогда и только тогда, когда существует строка категорий $\Gamma = \alpha_1^{\theta_1} \dots \alpha_n^{\theta_n}$, где $\alpha_i^{\theta_i} \in \delta(w_i)$, такая, что:*

1. $\alpha_1 \dots \alpha_n \vdash^* [S]$,

2. $\theta_1 \dots \theta_n$ сбалансирован.

Эта теорема была доказана для оКГЗ (без функции π) в статьях [7, 8]. Это доказательство без существенных изменений остаётся справедливым и для класса ммКГЗ.

Известно, что для кс-грамматик существуют различные специальные формы (нормальная форма Хомского, нормальная форма Грейбах(см. [1, 3])), упрощающие анализ порождаемых ими языков.

Определение 6. *Будем называть ммКГЗ $G =< W, \mathbf{C}, S, \delta, \pi >$ грамматикой в нормальной форме, если все её категории имеют один из следующих видов: $[X]^\theta$, $[X/Y]^\theta$, $[X/Y/Z]^\theta$, где X, Y, Z — элементарные категории, θ — потенциал.*

Для ммКГЗ справедливо следующее утверждение.

Теорема (о нормальной форме) *Для любой ммКГЗ $G =< W, \mathbf{C}, S, \delta, \pi >$ существует ммКГЗ $G' =< W, \mathbf{C}', S, \delta', \pi >$ в нормальной форме такая, что $L(G) = L(G')$.*

Доказательство этой теоремы для ммКГЗ аналогично доказательству о нормальной форме из работы [5]. В статье [8] доказано, что для любой оКГЗ G существует оКГЗ G' такая, что $L(G) = L(G')$ и G' не содержит главной категории в списках зависимостей. Если применить предложенное построение к грамматике в нормальной форме, то в результате снова получится грамматика в нормальной форме. Поэтому можно считать, что ммКГЗ в нормальной форме не содержат S в списках зависимостей.

Пусть $G =< W, \mathbf{C}, S, \delta, \pi >$ — ммКГЗ в нормальной форме. Пусть $\Gamma = \gamma_1 \dots \gamma_n$ — строка категорий такая, что $\Gamma \vdash^* [S]$. Тогда γ_1 имеет один из трёх видов: $[S]^\theta$, $[S/A]^\theta$ или $[S/A/B]^\theta$. Это следует из того, что в ммКГЗ в нормальной форме самая левая категория не может сократиться с другими категориями, так как левые списки зависимостей пустые.

3. Свойства замкнутости

В статьях [8] и [5] доказано, что класс языков, задаваемых оКГЗ, замкнут относительно операций объединения, конкатенации, пересечения с регулярными языками, неукорачивающего гомоморфизма и обращения гомоморфизма. Эти доказательства непосредственно переносятся на класс ммКГЗ-языков. В этом разделе мы установим, что класс ммКГЗ-языков замкнут относительно итерации. Отметим, что вопрос о замкнутости класса оКГЗ-языков относительно итерации остаётся открытым.

Теорема 1. *Пусть L является ммКГЗ-языком. Тогда L^* также является ммКГЗ-языком.*

Доказательство. Пусть язык L задаётся грамматикой $G =< W, \mathbf{C}, S, \delta, \pi >$. Будем считать, что G — грамматика в нормальной форме, не содержащая S в списках зависимостей. Построим индукцией по i последовательность грамматик $G_i =< W, \mathbf{C}_i, S, \delta_i, \pi >$.
Базис индукции. $i = 1$. Если в G у слова w есть категория $[S/A/\alpha]^\theta$, то в G_1 добавим этому слову категорию $[S/A'/\alpha]^\theta$.
Индукционный шаг. Пусть в грамматике G_i есть новые категории $A'_1, \ldots, A'_k$. Если в G_i у слова w была категория $[A_j/\alpha]^\theta$, $1 \leqslant j \leqslant k$, то в G_{i+1} добавим для w категорию$[A'_j/\alpha]^\theta$. После этого для каждой категории $[A'_j/A/\alpha]^\theta$ добавим категорию $[A'_j/A'/\alpha]^\theta$.
Этот процесс остановится не более чем через $|\mathbf{C}|$ шагов. Пусть последняя грамматика —
$G_m =< W, \mathbf{C}_m, S, \delta_m, \pi >$. Пусть $b \notin \mathbf{C}_m$ — новая категория. Построим грамматику $G'_m =< W, \mathbf{C}_m \cup \{\ b\ \}, S, \delta'_m, \pi' >$ следующим образом. Категории вида $[A']^\theta$ заменим на $[A']^{\theta \searrow b}$, категории вида $[S/A'/\alpha]^\theta$ заменим на $[S/A'/\alpha]^{\nearrow b\theta}$. Если в G_m есть категории вида $[S]^\theta$, то их заменим на $[S]^{\nearrow b\theta\searrow b}$. Наконец, $\pi'(A) = \pi(A) \cup \{\ b\ \}$ для всех $A \in \mathbf{C}_m$, $\pi'(b) = \mathbf{C}_m$.
Теперь можно построить грамматику, задающую итерацию исходного языка: $G' =< W, \mathbf{C}_m \cup \{\ b\ \}, S_0, \delta', \pi' >$. Если в G'_m есть категория $[S/\alpha]^\theta$, то в G' добавим $[S_0/S^*/\alpha]^\theta$

Лемма 1. *Пусть $\gamma_1 \ldots \gamma_n$ — строка категорий из G'_m. Если $\gamma_1 \ldots \gamma_n \vdash^* [A]^\theta$, то в $\gamma_1, \ldots, \gamma_n$ нет новых категорий вида X'.*

Доказательство индукцией по n.
Базис индукции. $n = 1$. Тогда $\gamma_1 = [A]^\theta$ и категорий вида X' нет.
Индукционный шаг. Пусть $\gamma_1 \ldots \gamma_{n+1} \vdash^* [A]^\theta$.
1) $\gamma_1 = [A/B]^{\theta_1}$. Тогда $\gamma_2 \ldots \gamma_{n+1} \vdash^* [B]^{\theta_2}$. По индукционному предположению категорий вида X' нет.
2) $\gamma_1 = [A/B/C]^{\theta_1}$. Тогда $\gamma_2 \ldots \gamma_r \vdash^* [C]^{\theta_2}$, $\gamma_{r+1} \ldots \gamma_{n+1} \vdash^*_{G'_m} [B]^{\theta_3}$. По индукционному предположению категорий вида X' нет. □

Лемма 2. *Пусть $\gamma_1 \ldots \gamma_n$ — строка категорий из G'_m. Пусть $\gamma_1 \ldots \gamma_n \vdash^* [A']^\theta$. Тогда $\gamma_n = [X']^{\theta' \searrow b}$, а потенциалы остальных категорий не содержат b.*

Доказательство индукцией по n.
Базис индукции. $n = 1$. Тогда $\gamma_1 = [X']^\theta$ и по построению $\theta = \theta' \searrow b$.
Индукционный шаг. Пусть $\gamma_1 \ldots \gamma_{n+1} \vdash^* [A']^\theta$.
1) $\gamma_1 = [A'/B']^{\theta_1}$. Тогда $\gamma_2 \ldots \gamma_n \vdash^* [B']^{\theta_2}$. По индукционному предположению $\gamma_{n+1} = [X']^{\theta' \searrow b}$ и потенциалы $\gamma_2, \ldots, \gamma_{n+1}$ не содержат b. θ_1 не содержит b по построению.
2) $\gamma_1 = [A'/B'/C]^{\theta_1}$. Тогда $\gamma_2 \ldots \gamma_r \vdash^* [C]^{\theta_2}$, $\gamma_{r+1} \ldots \gamma_{n+1} \vdash^*_{G'_m} [B']^{\theta_3}$. По индукционному предположению $\gamma_{n+1} = [X']^{\theta' \searrow b}$ и потенциалы $\gamma_{r+1}, \ldots, \gamma_{n+1}$ не содержат b. По лемме 1 $\gamma_2, \ldots, \gamma_r$ не содержат категорий вида Y', а значит по построению их потенциалы не содержат b. θ_1 не содержит b также по построению. □

Лемма 3. *Пусть $\Gamma = \gamma_1 \ldots \gamma_n \in \delta'_m(w)$, $\Gamma \vdash^* [A]^\theta$ или $\Gamma \vdash^* [A']^{\theta \searrow b}$ $(A \neq S)$. Тогда существует строка категорий $\Gamma' = \gamma'_1 \ldots \gamma'_n \in \delta(w)$ такая, что $\Gamma' \vdash^* [A]^\theta$.*

Доказательство. Если $\gamma_1 \dots \gamma_n \vdash^* [A]^\theta$, то по лемме 1 γ_i не содержат категорий X'. Это значит, что все категории γ_i были в исходной грамматике G. Поэтому можно взять $\Gamma' = \Gamma$. Доказательство для второго случая проведём индукцией по n.
Базис индукции. $n = 1$.
$\gamma_1 = [A']^{\theta \searrow b}$. Можно взять $\gamma_1' = [A]^\theta$.
Индукционный шаг. Пусть $\gamma_1 \dots \gamma_{n+1} \vdash^* [A']^{\theta \searrow b}$.
1) $\gamma_1 = [A'/B']^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_{n+1} \vdash^* [B']^{\theta_2 \searrow b}$. По индукционному предположению существует строка категорий $\gamma_2' \dots \gamma_{n+1}'$ такая, что $\gamma_2' \dots \gamma_{n+1}' \vdash^* [B]^{\theta_2}$. Пусть $\gamma_1' = [A/B]^{\theta_1}$. Тогда $\gamma_1' \dots \gamma_{n+1}' \vdash^* [A]^{\theta_1\theta_2} = [A]^\theta$.
2) $\gamma_1 = [A'/B'/C']^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_r \vdash^* [C]^{\theta_2}$, $\gamma_{r+1} \dots \gamma_{n+1} \vdash^* [B']^{\theta_3 \searrow b}$. Существуют строки категорий $\gamma_2' \dots \gamma_r'$ и $\gamma_{r+1}' \dots \gamma_{n+1}'$ такие, что $\gamma_2' \dots \gamma_r' \vdash^* [C]^{\theta_2}$, $\gamma_{r+1}' \dots \gamma_{n+1}' \vdash^* [B]^{\theta_3}$. Пусть $\gamma_1' = [A/B/C]^{\theta_1}$. Тогда $\gamma_1' \dots \gamma_{n+1}' \vdash^* [A]^{\theta_1\theta_2\theta_3} = [A]^\theta$.
$\square$

Лемма 4. *Пусть* $\Gamma = \gamma_1 \dots \gamma_n \in \delta(w)$, $\Gamma \vdash^* [A]^\theta (A \neq S)$. *Тогда если* $A' \in \mathbf{C}_m$, *то существует* $\Gamma' = \gamma_1' \dots \gamma_n' \in \delta_m'(w)$ *такая, что* $\Gamma' \vdash^* [A']^{\theta \searrow b}$.

Доказательство индукцией по n.
Базис индукции. $n = 1$. $\gamma_1 = [A]^\theta$. Тогда $\gamma_1' = [A']^{\theta \searrow b}$.
Индукционный шаг. Пусть $\Gamma = \gamma_1 \dots \gamma_{n+1}$.
1) $\gamma_1 = [A/B]^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_{n+1} \vdash^* [B]^{\theta_2}$. Так как была добавлена категория A', то была добавлена и B'. По индукционному предположению существует строка категорий $\gamma_2' \dots \gamma_{n+1}'$ такая, что $\gamma_2' \dots \gamma_{n+1}' \vdash^* [B']^{\theta_2 \searrow b}$. Пусть $\gamma_1' = [A'/B']^{\theta_1}$. Имеем $\gamma_1' \dots \gamma_{n+1}' \vdash^* [A'/B']^{\theta_1}[B']^{\theta_2 \searrow b} \vdash [A']^{\theta_1\theta_2 \searrow b} = [A']^{\theta \searrow b}$.
2) $\gamma_1 = [A/B/C]^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_r \vdash^* [C]^{\theta_2}$, $\gamma_{r+1} \dots \gamma_{n+1} \vdash^* [B]^{\theta_3}$. Все категории из G, не содержащие S, остались и в G_m'. Также G_m' содержит B'. Поэтому существует строка категорий $\gamma_2' \dots \gamma_r'$ такая, что $\gamma_2' \dots \gamma_r' \vdash^* [C]^{\theta_2}$ (достаточно взять $\gamma_i' = \gamma_i$) и строка категорий $\gamma_{r+1}' \dots \gamma_{n+1}'$ такая, что $\gamma_{r+1}' \dots \gamma_{n+1}' \vdash^* [B']^{\theta_3 \searrow b}$. Возьмём $\gamma_1' = [A'/B'/C]^{\theta_1}$. Тогда $\gamma_1 \dots \gamma_{n+1} \vdash^* [A'/B'/C]^{\theta_1}[C]^{\theta_2}[B']^{\theta_3 \searrow b} \vdash^* [A']^{\theta_1\theta_2\theta_3 \searrow b} = [A']^{\theta \searrow b}$.
$\square$

Лемма 5. $L(G) = L(G_m')$

Доказательство.
[$\Rightarrow$] Пусть $w = w_1 \dots w_n \in L(G)$. По определению существует строка категорий $\gamma_1 \dots \gamma_n \in \delta(w)$ такая, что $\gamma_1 \dots \gamma_n \vdash^* [S]$. Нужно построить строку категорий Γ' в грамматике G_m' такую, что $\Gamma' \vdash^* [S]$. Для γ_1 возможны три случая.
1) $\gamma_1 = [S]^\theta$. В этом случае $w = w_1$. Присвоим w_1 категорию $[S]^{\nearrow b\theta \searrow b}$. Так как θ сбалансирован, то $[S]^{\nearrow b\theta \searrow b} \vdash^* [S]^{\nearrow b \searrow b} \vdash [S]$.
2) $\gamma_1 = [S/A]^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_n \vdash^* [A]^{\theta_2}$ и потенциал $\theta_1\theta_2$ сбалансирован. В грамматике G_m' у слова w_1 есть категория $\gamma_1' = [S/A']^{\nearrow b\theta_1}$. По лемме 4 существует строка категорий $\gamma_2' \dots \gamma_n' \in \delta_m'(w_2 \dots w_n)$ такая, что $\gamma_2' \dots \gamma_n' \vdash^* [A']^{\theta_2 \searrow b}$. Тогда $\gamma_1' \dots \gamma_n' \vdash^* [S/A']^{\nearrow b\theta_1}[A']^{\theta_2 \searrow b} \vdash [S]^{\nearrow b\theta_1\theta_2 \searrow b} \vdash^* [S]^{\nearrow b \searrow b} \vdash [S]$.
3) $\gamma_1 = [S/A/B]^{\theta_1}$. Тогда $\gamma_2 \dots \gamma_r \vdash^* [B]^{\theta_2}$, $\gamma_{r+1} \dots \gamma_n \vdash^* [A]^{\theta_3}$ и потенциал $\theta_1\theta_2\theta_3$ сбалансирован. В грамматике G_m' у слова w_1 есть категория $\gamma_1' = [S/A'/B]^{\nearrow b\theta_1}$. Так как все категории из G, не содержащие S сохранились, то $\gamma_2 \dots \gamma_r \in \delta_m'(w_2 \dots w_r)$. По лемме 4 существует строка категорий $\gamma_{r+1}' \dots \gamma_n' \in \delta_m'(w_{r+1} \dots w_n)$ такая, что $\gamma_{r+1}' \dots \gamma_n' \vdash^* [A']^{\theta_3 \searrow b}$. Тогда $\gamma_1' \dots \gamma_n' \vdash^* [S/A'/B]^{\nearrow b\theta_1}[B]^{\theta_2}[A']^{\theta_3 \searrow b} \vdash^*$

$[S]^{\nearrow b\theta_1\theta_2\theta_3\searrow b} \vdash^* [S]^{\nearrow b\searrow b} \vdash [S]$.
Во всех случаях $w \in L(G'_m)$.
[$\Leftarrow$] Пусть $w = w_1 \dots w_n \in L(G'_m)$. По определению существует строка категорий $\gamma_1 \dots \gamma_n \in \delta'_m(w)$ такая, что $\gamma_1 \dots \gamma_n \vdash^* [S]$. Для γ_1 возможны три случая.
1) $\gamma_1 = [S]^{\nearrow b\theta\searrow b}$. В этом случае $w = w_1$. Присвоим w_1 категорию $[S]^\theta$. Потенциал θ сбалансирован, поэтому $[S]^\theta \vdash^* [S]$.
2) $\gamma_1 = [S/A']^{\nearrow b\theta_1}$. Тогда $\gamma_2 \dots \gamma_n \vdash^* [A']^{\theta_2\searrow b}$, причём потенциал $\theta_1\theta_2$ сбалансирован. По лемме 3 существует строка категорий $\gamma'_2 \dots \gamma'_n \in \delta(w_2 \dots w_n)$ такая, что $\gamma'_2 \dots \gamma'_n \vdash^* [A]^{\theta_2}$. Слову w_1 припишем категорию $\gamma'_1 = [S/A]^{\theta_1}$. Имеем $\gamma'_1 \dots \gamma'_n \vdash^* [S/A]^{\theta_1}[A]^{\theta_2} \vdash [S]^{\theta_1\theta_2} \vdash^* [S]$.
3) $\gamma_1 = [S/A'/B]^{\nearrow b\theta_1}$. Тогда $\gamma_2 \dots \gamma_r \vdash^* [B]^{\theta_2}$, $\gamma_{r+1} \dots \gamma_n \vdash^* [A']^{\theta_3\searrow b}$, причём потенциал $\theta_1\theta_2\theta_3$ сбалансирован. Возьмём $\gamma'_2 \dots \gamma'_r = \gamma_2 \dots \gamma_r$. По лемме 3 существует строка категорий $\gamma'_{r+1} \dots \gamma'_n \in \delta(w_{r+1} \dots w_n)$ такая, что $\gamma'_{r+1} \dots \gamma'_n \vdash^* [A]^{\theta_3}$. Слову w_1 припишем категорию $\gamma'_1 = [S/A/B]^{\theta_1}$. Имеем $\gamma'_1 \dots \gamma'_n \vdash^* [S/A/B]^{\theta_1}[B]^{\theta_2}[A]^{\theta_3} \vdash^* [S]^{\theta_1\theta_2\theta_3} \vdash^* [S]$.
Во всех случаях $w \in L(G)$. $\square$

Ввиду этой леммы для доказательства теоремы достаточно установить, что $L(G'_m)^* = L(G')$.
[$\Rightarrow$] $L(G'_m)^* \subseteq L(G')$
Пусть $w \in L(G'_m)^*$, т. е. $w = w_1 \dots w_n$, $w_i \in L(G'_m)$. Это означает, что для каждого i существует строка категорий $\Gamma_i \in \delta'_m(w_i)$ такая, что $\Gamma_i \vdash^* [S]$. Первая категория в Γ_1 имеет вид $[S/\alpha]^{\nearrow b\theta_1}$. Заменим её на категорию $[S_0/S^*/\alpha]^{\nearrow b\theta_1}$. Получится новая строка категорий $\Gamma'_1 \in \delta'(w_1)$, $\Gamma'_1 \vdash^* [S_0/S^*]$. Припишем слову w строку категорий $\Gamma'_1\Gamma_2 \dots \Gamma_n$. Имеем $\Gamma'_1\Gamma_2 \dots \Gamma_n \vdash^* [S_0/S^*][S] \dots [S] \vdash^* [S_0]$, т. е. $w \in L(G')$.
[$\Leftarrow$] $L(G') \subseteq L(G'_m)^*$
Пусть $w \in L(G')$. Это значит, что существует строка категорий $\Gamma \in \delta'(w)$ такая, что $\Gamma \vdash^* [S_0]$. После нескольких сокращений получится строка $[S_0/S^*]^{\theta_1}[S]^{\theta_2} \dots [S]^{\theta_n}$, причём сокращений в потенциалах пока не было и категории S не сокращались. Пусть $\Gamma = \Gamma_1 \dots \Gamma_n$, так что $\Gamma_1 \vdash^* [S_0/S^*]^{\theta_1}$, $\Gamma_i \vdash^* [S]^{\theta_i}$ при $i \geqslant 2$, и пусть строка Γ_i приписана подслову w_i. Для всех i (включая $i = 1$) в грамматике G'_m можно выполнить сокращения $\Gamma_i \vdash^* [S]^{\theta_i}$. Все потенциалы θ_i сбалансированы. Действительно, если $|w_i| > 1$, то $\Gamma_i = [S/\alpha]^{\nearrow b\theta_i^1}[\alpha_2]^{\theta_i^2} \dots [\alpha_{k-1}]^{\theta_i^{k-1}}[A']^{\theta_i^k\searrow b}$ (то, что последняя категория имеет вид $[A']^{\theta_i^k\searrow b}$, следует из леммы 2 и из построения G'_m). Из лемм 1 и 2 следует, что потенциал $\theta_i^1 \dots \theta_i^k$ не содержит стрелок b. Поэтому он сбалансирован, иначе крайние стрелки $\nearrow b$ и $\searrow b$ не пустили бы оставшиеся валентности. А тогда сбалансирован и потенциал θ_i. Если же $|w_i| = 1$, то $\theta_i = \nearrow b\theta'_i \searrow b$ и θ_i также сбалансирован. Мы получили, что $\Gamma_i \vdash^* [S]$, т. е. $w_i \in L(G'_m)$. А это и означает, что $w \in L(G'_m)^*$. $\square$

В статье [11] было введено понятие *абстрактного семейства языков* (АСЯ). Класс языков является АСЯ, если он замкнут относительно операций объединения, конкатенации, пересечения с регулярными языками, неукорачивающего гомоморфизма, обращения гомоморфизма и итерации. Класс ммКГЗ-языков замкнут относительно этих шести операция, поэтому справедливо следующее утверждение.

Теорема 2. *Класс $\mathcal{L}(mmCDG)$ является абстрактным семейством языков.*

Известно, что класс кс-языков является АСЯ, но он не замкнут относительно пересечения. Для класса же ммКГЗ-языков справедлива следующая теорема.

Теорема 3. *Класс ммКГЗ-языков замкнут относительно пересечения.*

Доказательство. Пусть $G_1 =< W, \mathbf{C}_1, S_1, \delta_1, \pi_1 >$ и $G_2 =< W, \mathbf{C}_2, S_2, \delta_2, \pi_2 >$ — две ммКГЗ. Пусть $\mathbf{C}_2^{loc} \subseteq \mathbf{C}_2$ — множество элементарных категорий, использованных в локальных частях, и пусть $\mathbf{C}_2^{val} \subseteq \mathbf{C}_2$ — множество элементарных категорий, использованных в валентностях. Мы можем предполагать, что $\mathbf{C}_1 \cap \mathbf{C}_2 = \varnothing$, $\mathbf{C}_2^{loc} \cap \mathbf{C}_2^{val} = \varnothing$, S_1 и S_2 не встречаются в списках зависимостей. Теперь мы построим грамматику $G =< W, \mathbf{C}_1 \cup \mathbf{C}_2, S_1, \delta, \pi >$ такую, что $L(G) = L(G_1) \cap L(G_2)$. Для простоты будем считать, что G_2 находится нормальной форме. Мы преобразуем локальные категории G_2 в потенциалы, которые моделируют локальные сокращения. Сначала определим функцию p, которая описывает это преобразование.
$p([A]) = \searrow A$, если $A \neq S_2$
$p([A/B]) = \searrow A \nearrow B$, если $A \neq S_2$
$p([A/B/C]) = \searrow A \nearrow B \nearrow C$, если $A \neq S_2$
$p([S_2]) = \varepsilon$
$p([S_2/B]) = \nearrow B$
$p([S_2/B/C]) = \nearrow B \nearrow C$
Теперь определим словарь δ.
1) Пусть $\alpha^{\theta} \in \delta_1(w)$, $\beta^{\theta'} \in \delta_2(w)$ и α, β не содержат S_1, S_2. Тогда $\alpha^{\theta\theta' p(\beta)} \in \delta(w)$.
2) Пусть $\alpha^{\theta} \in \delta_1(w)$, $\beta^{\theta'} \in \delta_2(w)$, где $\alpha = [S_1/\alpha']$, $\beta = [S_2/\beta']$. Тогда $\alpha^{\theta\theta' p(\beta)} \in \delta(w)$.
Ограничения будут следующими.
$\pi(X) = \pi_1(X)$ для $X \in \mathbf{C}_1$
$\pi(X) = \pi_2(X)$ для $X \in \mathbf{C}_2^{val}$
$\pi(X) = \mathbf{C}_2^{loc}$ для $X \in \mathbf{C}_2^{loc}$

Лемма 6. *Пусть $\gamma_1, \ldots, \gamma_n$ — строка категорий, содержащих элементарные категории только из $\mathbf{C}_2^{loc}$, в которой не встречается S_2. Тогда $\gamma_1 \ldots \gamma_n \vdash^* [A]$ тогда и только тогда, когда $p(\gamma_1) \ldots p(\gamma_n) \vdash^* \searrow A$.*

Доказательство индукцией по n.
$[\Rightarrow]$ Пусть $\gamma_1 \ldots \gamma_n \vdash^* [A]$.
Базис индукции. $n = 1$. Тогда $\gamma_1 = [A]$, и $p(\gamma_1) = \searrow A$.
Индукционный шаг. Пусть $\gamma_1 \ldots \gamma_{n+1} \vdash^* [A]$.
1) Случай $\gamma_1 = [A]$ невозможен, потому что грамматика находится в нормальной форме.
2) $\gamma_1 = [A/B]$, $\gamma_2 \ldots \gamma_{n+1} \vdash^* [B]$. Тогда по индукционному предположению $p(\gamma_2) \ldots p(\gamma_{n+1}) \vdash^* \searrow B$, и по определению p $p(\gamma_1) = \searrow A \nearrow B$. Тогда $p(\gamma_1) \ldots p(\gamma_{n+1}) \vdash^* \searrow A \nearrow B \searrow B \vdash^* \searrow A$.
3) $\gamma_1 = [A/B/C]$, $\gamma_2 \ldots \gamma_k \vdash^* [C]$, $\gamma_{k+1} \ldots \gamma_{n+1} \vdash^* [B]$ для некоторого k. Этот случай аналогичен предыдущему. $\gamma_1 \ldots \gamma_{n+1} \vdash^* \searrow A \nearrow B \nearrow C \searrow C \searrow B \vdash^* \searrow A$.
$[\Leftarrow]$ Пусть $p(\gamma_1) \ldots p(\gamma_n) \vdash^* \searrow A$.
Базис индукции. $n = 1$. $p(\gamma_1) = \searrow A$. Тогда по определению $\gamma_1 = [A]$.
Индукционный шаг. Пусть $p(\gamma_1) \ldots p(\gamma_{n+1}) \vdash^* \searrow A$. Так как γ_i не содержит S_2, то каждое $p(\gamma_i)$ начинается с некоторой отрицательной валентности. В частности, γ_1 начинается с $\searrow A$.
1) $p(\gamma_1) = \searrow A$, $p(\gamma_2) \ldots p(\gamma_{n+1}) \vdash^* \varepsilon$. Этот случай невозможен, так как $p(\gamma_2)$ начинается с отрицательной валентности.

2) $p(\gamma_1) = \searrow A \nearrow B$, $p(\gamma_2)\dots p(\gamma_{n+1}) \vdash^* \searrow B$. Тогда по индукционному предположению $\gamma_2 \dots \gamma_{n+1} \vdash^* [B]$, и по определению $\gamma_1 = [A/B]$. Тогда $\gamma_1 \dots \gamma_{n+1} \vdash^* [A]$.
3) $p(\gamma_1) = \searrow A \nearrow B \nearrow C$, $p(\gamma_2)\dots p(\gamma_{n+1}) \vdash^* \searrow C \searrow B$. Валентности расположены в порядке C, B из-за ограничений π. Так как каждое γ_i начинается с отрицательной валентности, то никакую подстроку $p(\gamma_2)\dots p(\gamma_{n+1})$ нельзя сократить до ε. Поэтому существует такое k, что $p(\gamma_2)\dots p(\gamma_k) \vdash^* \searrow C$, $p(\gamma_{k+1})\dots p(\gamma_{n+1}) \vdash^* \searrow B$. Поэтому $\gamma_1 \dots \gamma_{n+1} \vdash^* [A/B/C][C][B] \vdash^* [A]$. □

Лемма 7. *Пусть $\gamma_1, \dots, \gamma_n$ — локальные категории, содержащие элементарные категории только из $\mathbf{C}_2^{loc}$, такие что $\gamma_2, \dots, \gamma_n$ не содержат S_2, и γ_1 содержит S_2. Тогда $\gamma_1 \dots \gamma_n \vdash^* [S_2]$ тогда и только тогда, когда $p(\gamma_1)\dots p(\gamma_n) \vdash^* \varepsilon$.*

Доказательство.
1) $\gamma_1 = [S_2]$ тогда и только тогда, когда $p(\gamma_1) = \varepsilon$.
2) $\gamma_1 = [S_2/A]$. Тогда $\gamma_1 \dots \gamma_n \vdash^* [S_2]$ тогда и только тогда, когда $\gamma_2 \dots \gamma_n \vdash^* [A]$. Но по лемме 6 это выполняется тогда и только тогда, когда $p(\gamma_2)\dots p(\gamma_n) \vdash^* \searrow A$. Это в свою очередь выполняется тогда и только тогда, когда $p(\gamma_1)\dots p(\gamma_n) \vdash^* \varepsilon$.
3) $\gamma_1 = [S_2/A/B]$. Тогда $\gamma_1 \dots \gamma_n \vdash^* [S_2]$ тогда и только тогда, когда для некоторого k $\gamma_2 \dots \gamma_k \vdash^* [B]$, $\gamma_{k+1} \dots \gamma_n \vdash^* [A]$. По лемме 6 это выполняется тогда и только тогда, когда $p(\gamma_2)\dots p(\gamma_k) \vdash^* \searrow B$, $p(\gamma_{k+1})\dots p(\gamma_n) \vdash^* \searrow A$. Это выполняется тогда и только тогда, когда $p(\gamma_1)\dots p(\gamma_n) \vdash^* \varepsilon$ (как и в доказательстве леммы 6 мы используем ограничения π и тот факт, что никакую подстроку нельзя сократить до ε). □

Теперь мы можем доказать, что $L(G) = L(G_1) \cap L(G_2)$.
[$\Rightarrow$] Пусть $w \in L(G)$. Тогда существует строка категорий $\alpha_1^{\theta_1} \dots \alpha_n^{\theta_n} \in \delta(w)$ такая, что $\alpha_1 \dots \alpha_n \vdash^* [S_1]$ и $\theta_1 \dots \theta_n$ сбалансирован. Каждое θ_i состоит из трёх частей: $\theta_i = \theta_i' \theta_i'' \theta_i'''$. Здесь θ_i' — потенциал категории α_i в G_1, θ_i'' — потенциал некоторой категории из G_2, а θ_i''' представляет локальную часть этой категории. $\theta_1' \dots \theta_n'$ сбалансирован, так как $\mathbf{C}_1 \cap \mathbf{C}_2 = \varnothing$ и $\pi(X)$ для $X \in \mathbf{C}_1$ не содержат категорий из $\mathbf{C}_2$. Это значит, что $w \in L(G_1)$. $\theta_1'' \dots \theta_n''$ и $\theta_1''' \dots \theta_n'''$ также сбалансированы. Пусть β_i — локальные категории, такие что $p(\beta_i) = \theta_i'''$. G_1 находится в нормальной форме, поэтому только у α_1 есть S_1. Но это значит, что только θ_1''' не содержит отрицательных валентностей. Тогда по лемме 7 $\beta_1 \dots \beta_n \vdash^* [S_2]$, т.е. $w \in L(G_2)$.
[$\Leftarrow$] Пусть $w \in L(G_1), w \in L(G_2)$. Тогда существуют строки категорий $\alpha_1^{\theta_1} \dots \alpha_n^{\theta_n} \in \delta_1(w)$ и $\beta_1^{\eta_1} \dots \beta_n^{\eta_n} \in \delta_2(w)$ такие, что $\alpha_1 \dots \alpha_n \vdash^* [S_1]$, $\beta_1 \dots \beta_n \vdash^* [S_2]$, $\theta_1 \dots \theta_n$ и $\eta_1 \dots \eta_n$ сбалансированы. Присвоим следующую строку категорий слову w в грамматике G: $\alpha_1^{\theta_1 \eta_1 p(\beta_1)} \dots \alpha_n^{\theta_n \eta_n p(\beta_n)}$. Нужно только доказать, что потенциал этой строки сбалансирован. Сначала мы можем сократить $\theta_1 \dots \theta_n$, так как валентности в остальной части потенциала не встречаются в ограничениях для $\mathbf{C}_1$. Потом мы можем сократить $\eta_1 \dots \eta_n$. G_2 находится в нормальной форме, поэтому только в β_1 есть S_2. Тогда по лемме 7 $p(\beta_1)\dots p(\beta_n)$ сбалансирован. Это значит, что весь потенциал сбалансирован и $w \in L(G)$. □

В заключение рассмотрим вопрос о замкнутости класса ммКГЗ-языков относительно проекции.

Теорема 4. *Класс ммКГЗ-языков не замкнут относительно проекции.*

Доказательство. Пусть L — рекурсивно перечислимый, но нерекурсивный язык. Существуют два кс-языка L_1, L_2 и гомоморфизм φ такие, что $L = \varphi(L_1 \cap L_2)$

([10]). Но и L_1, и L_2 — ммКГЗ-языки, поэтому $L_1 \cap L_2$ — тоже ммКГЗ-язык. Так как любой гомоморфизм можно представить как произведение проекции и неукорачивающего гомоморфизма и класс ммКГЗ-языков замкнут относительно неукорачивающего гомоморфизма, то из замкнутости класса ммКГЗ-языков относительно проекции следовала бы его замкнутость и относительно произвольных гомоморфизмов. Но тогда L был бы ммКГЗ-языком. А это невозможно, потому что все ммКГЗ-языки рекурсивны (проблема принадлежности для ммКГЗ лежит в классе NP, см. раздел 5). □

4. Неполулинейность

Пусть $L \subseteq \mathbb{N}^n$ — произвольное множество векторов. Множество L называется линейным, если существуют векторы $c, x_1, \ldots, x_m$ такие, что $L = \{c + k_1x_1 + \cdots + k_mx_m \mid k_i \in \mathbb{N}\}$. Подмножество множества $\mathbb{N}^n$ называется полулинейным, если оно является объединением конечного числа линейных множеств. Пусть $W = \{w_1, \ldots, w_n\}$ — конечное множество символов. Обозначим через ψ отображение множества W^* во множество $\mathbb{N}^n$, определяемое следующим образом: $\psi(s) = (|s|_{w_1}, \ldots, |s|_{w_n})$, где $|s|_{w_i}$ — число вхождений символа w_i в слово s. Если $L \subseteq W^*$, то $\psi(L) = \{\psi(s) \mid s \in L\}$. Язык L называется полулинейным, если множество $\psi(L)$ полулинейно. Имеет место следующая теорема (см. [2]).
Теорема Парика *Для каждого кс-языка L множество $\psi(L)$ является полулинейным.*

Одним из известных свойств полулинейных языков, является то, что длины входящих в них слов содержат бесконечные арифметические прогрессии.

В этом разделе мы доказываем, что класс ммКГЗ-языков содержит неполулинейные языки. Примером является язык $L = \{\, 101001 \ldots 10^{2^n} \mid n = 3, 5, 7, \ldots \}$. Множество длин слов этого языка не содержит бесконечных арифметических прогрессий, поэтому L — неполулинейный язык.

Теорема 5. $L = \{\, 101001 \ldots 10^{2^n} \mid n = 3, 5, 7, \ldots \}$ *является ммКГЗ-языком.*

Доказательство. L порождается следующей ммКГЗ G.
$0 \mapsto [X_1/X_2], [X_3/X_4]^{\nearrow A \nearrow A}, [X_4/A]^{\nearrow A \nearrow A}, [B/D/D/B]^{\searrow A}, [D]^{\nearrow A \nearrow A}, [C/E/E/C]^{\searrow A}, [E]$
$1 \mapsto [S/X_1], [X_2/X_3]^{\nearrow B}, [A/A/B], [A/C], [B]^{\searrow B \nearrow B}, [C]^{\searrow B}$
$\pi(A) = \{\, B \,\}, \pi(B) = \{\, A \,\}$

С помощью локальных категорий G проверяет, что каждый чётный блок нулей в два раза длиннее предыдущего нечётного блока. Дальние связи типа A используются, чтобы проверить, что каждый нечётный блок нулей в два раза длиннее предыдущего чётного блока. Чтобы гарантировать, что все стрелки типа A будут пойманы ближайшим блоком нулей, мы используем стрелки типа B. Они соединяют единицы, не позволяя стрелкам типа A попасть не в тот блок. В начале слова ещё нет выпущенных стрелок, которые можно было бы поймать, поэтому мы используем специальные категории (они содержат X). В конце слова не нужно выпускать новые стрелки направо, поэтому мы используем категории, содержащие C и E. Их единственное отличие от категорий с B и D состоит в том, что они только ловят стрелки.

Лемма 8. *Пусть* $w = 10^{i_1}10^{i_2}\dots 10^{i_{2k}}w'' \in L(G)$, $w' = 10^{i_1}10^{i_2}\dots 10^{i_{2k}}$. *Тогда:*
1) $i_j = 2^{j-1}$ *для* $1 \leqslant j \leqslant 2k$.
2) Если $w'' \neq \varepsilon$, *то* $\theta(w') \vdash^* \nearrow B(\nearrow A)^{2^{2k}}$.
3) Локальные категории строки $10^{i_{2j-1}}10^{i_{2j}}$ *можно сократить до* $[A/A]$ *для* $j > 1$.

Доказательство индукцией по k.
Базис индукции. $k = 1$. Единственное возможное начало имеет вид $[S/X_1][X_1/X_2][X_2/X_3]^{\nearrow B}$ $[X_3/X_4]^{\nearrow A \nearrow A}$ $[X_4/A]^{\nearrow A \nearrow A}$.
Индукционный шаг. По индукционному предположению $w = 10100\dots 10^{2^{2k-1}}w''$, $\theta(w') \vdash^* \nearrow B(\nearrow A)^{2^{2k}}$. Первой единице приписана либо категория $[A/A/B]$, либо $[A/C]$, так как категории предыдущего блока можно сократить до $[A/A]$.
1) Первой единице приписана категория $[A/A/B]$. Тогда следующие нули имеют категории $[B/D/D/B]^{\searrow A}$. Пусть длина этого блока нулей равна l. Тогда этому блоку соответствует строка категорий $([B/D/D/B]^{\searrow A})^l$. Мы докажем, что $l = 2^{2k}$. Предположим, что $l < 2^{2k}$. Тогда эта строка поймает l стрелок A. Но следующей категорией будет $[B]^{\searrow B \nearrow B}$. Тогда потенциал равен $\nearrow B(\nearrow A)^t \searrow B \nearrow B$, $t > 0$. Но это значит, что ограничивающая зависимость B не позволит стрелкам A пройти дальше, т.е. $w \notin L(G)$. Теперь предположим, что $l > 2^{2k}$. Тогда количество стрелок $\nearrow A$ меньше количества стрелок $\searrow A$, т.е. снова $w \notin L(G)$. Итак, $l = 2^{2k}$. У категории каждого нуля в этом блоке есть две категории D, и строка категорий всего блока содержит 2^{2k+1} категорий D. Это значит, что следующий блок должден содержать ровно 2^{2k+1} нулей. Каждый из них добавляет две стрелки $\nearrow A$, так что $\theta(w') \vdash^* \nearrow B(\nearrow A)^{2^{2k}}(\searrow A)^{2^{2k}} \searrow B \nearrow B(\nearrow A \nearrow A)^{2\cdot 2^{2k}} \vdash^* \nearrow B(\nearrow A)^{2^{2(k+1)}}$. Из присваивания категорий непосредственно следует, что строку категорий последних двух блоков можно сократить до $[A/A]$.
2) Первой единице приписана категория $[A/C]$. Доказательство повторяет предыдущий случай. Единственное отличие состоит в том, что последний блок не создаст дополнительных стрелок, так что потенциал будет сбалансированным. □

Из пункта 2 доказательства этой леммы следует, что $L(G) \subseteq L$. Наоборот, пусть слово $w = 10100\dots 10^{2^n} \in L$. Припишем слову следующую строку категорий Γ:

$$\dots[S/X_1][X_1/X_2][X_2/X_3]^{\nearrow B}[X_3/X_4]^{\nearrow A \nearrow A}[X_4/A]^{\nearrow A \nearrow A}\dots$$
$$\dots[A/A/B]([B/D/D/B]^{\searrow A})^4[B]^{\searrow B \nearrow B}([D]^{\nearrow A \nearrow A})^8\dots$$
$$\dots[A/A/B]([B/D/D/B]^{\searrow A})^{2^{2i}}[B]^{\searrow B \nearrow B}([D]^{\nearrow A \nearrow A})^{2^{2i+1}}\dots$$
$$[A/C]([C/E/E/C]^{\searrow A})^{2^{n-1}}[C]^{\searrow B}([E])^{2^n}$$

.

Рассмотрим подстроку $\Gamma_i = [A/A/B]([B/D/D/B]^{\searrow A})^{2^{2i}}[B]^{\searrow B \nearrow B}([D]^{\nearrow A \nearrow A})^{2^{2i+1}}$. Локальная категория B в этой подстроке может быть сокращена. После этого можно будет сократить две локальные категории D. Тогда от последней категории $[B/D/D/B]$ останется только $[B]$. Этот процесс будет продолжаться до тех пор, пока все локальные категории D не сократятся. Затем будет выполнено последнее сокращение $[A/A/B][B] \vdash [A/A]$. Ни одна валентность в Γ_i не может сократиться, поэтому $\theta(\Gamma_i) = (\searrow A)^{2^{2i}} \searrow B \nearrow B(\nearrow A)^{2^{2(i+1)}}$. Рассуждая таким же образом, получим, что локальные категории последних двух блоков сократятся до $[A]$, а потенциал будет равен $(\searrow A)^{2^{n-1}} \searrow B$. Исходящие стрелки $\nearrow A$ подстроки Γ_i сократятся с входящими стрелками $\searrow A$ подстроки Γ_{i+1}. После этого сокращения стрелки $\nearrow B$ и $\searrow B$ окажутся рядом и их можно будет сократить. Локальные же категории сократятся до $[A]$. Таким образом, получаем $\Gamma \vdash^* [S/X_1][X_1/X_2][X_2/X_3]^{\nearrow B}[X_3/X_4]^{\nearrow A \nearrow A}[X_4/A]^{\nearrow A \nearrow A}\ [A]^{\searrow A \searrow A \searrow A \searrow A \searrow B} \vdash^* [S]$. Итак, мы доказали, что $w \in L(G)$, т.е. $L \subseteq L(G)$. □

Следствие 1. *Класс ммКГЗ-языков содержит неполулинейные языки.*

Заметим, что вопрос о полулинейности КГЗ-языков. остаётся открытым.

5. Сложность проблемы принадлежности

В работе [7] было показано, что проблема определения по КГЗ G и слову w, принадлежит ли w языку $L(G)$, является NP-полной. Для класса ммКГЗ-языков этот результат может быть усилен. Прежде всего заметим, что проблема принадлежности для ммКГЗ лежит в классе NP. Определим следующее отношение $\prec$ на множестве дальних зависимостей (т.е. пар вида $a(d) = \nearrow d \searrow d$ или $a(d) = \swarrow d \nwarrow d$): $a(d_1) \prec_\pi a(d_2)$, если две зависимости пересекаются и $d_1 \in \pi(d_2)$. Это отношение означает, что зависимость $a(d_1)$ должна быть создана сокращением раньше, чем $a(d_2)$. Имеет место следующее свойство.

Лемма 9. *Потенциал θ сбалансирован тогда и только тогда, когда валентности из θ можно объединить в пары таким образом, что отношение $\prec_\pi$ не имеет циклов.*

Доказательство.
[⇒] Допустим, что отношение $\prec_\pi$ содержит цикл $a(d_1) \prec_\pi a(d_2) \prec_\pi \cdots \prec_\pi a(d_n) \prec_\pi a(d_1)$. Тогда пары, входящие в этот цикл, не сократятся. Действительно, пара $a(d_i)$ не может сократиться, так как ей мешает пара $a(d_{i-1})$ (пара $a(d_n)$ при $i = 1$). Поэтому потенциал θ не сбалансирован.
[⇐] Пусть отношение $\prec_\pi$ не содержит циклов. Проведём доказательство индукцией по длине потенциала.
Базис индукции. $|\theta| = 0$, т.е. $\theta = \varepsilon$. Этот потенциал сбалансирован.
Индукционный шаг. Так как множество пар валентностей конечно, то в нём есть минимальный элемент $a(d)$. Эту пару можно сократить, потому что ни одна валентность этому не препятствует. В результате получится потенциал θ', длина которого меньше длины θ. θ' не содержит циклов. По индукционному предположению он сбалансирован. Следовательно, сбалансирован и потенциал θ. □

Лемма 10. *Существует полиномиальный алгоритм для проверки сбалансированности потенциала.*

Доказательство. Пусть имеется потенциал $\theta = v_1 \dots v_k$, в котором используются n типов валентностей. Алгоритм использует n стеков $S_1 \dots S_n$.
1. $V := \varnothing$; $E := \varnothing$
2. ДЛЯ $i = 1$ ДО n ВЫПОЛНЯТЬ $S_i := \varnothing$ ВСЁ_ДЛЯ
3. ДЛЯ $i = 1$ ДО $|\theta|$ ВЫПОЛНЯТЬ

4. ЕСЛИ v_i — левая валентность типа j ТО
5. поместить i на вершину S_j
6. ИНАЧЕ ЕСЛИ v_i — правая валентность типа j И $S_j \neq \varnothing$ ТО
7. пусть l — элемент на вершине S_j
8. удалить вершину S_j
9. $V := V \cup \{ (i,l) \}$
10. ДЛЯ ВСЕХ $(p,q) \in V$ ВЫПОЛНЯТЬ
11. ЕСЛИ $p < i < l < q$ ТО $E := E \cup \{ ((i,l),(p,q)) \}$
12. ЕСЛИ $i < p < q < l$ ТО $E := E \cup \{ ((p,q),(i,l)) \}$
13. ЕСЛИ $p < i < q < l$ ИЛИ $i < p < l < q$ ТО $E := E \cup \{ ((i,l),(p,q)), ((p,q),(i,l)) \}$
14. ВСЁ_ДЛЯ
15. ИНАЧЕ
16. выдать "НЕТ"
17. ВСЁ_ЕСЛИ
18. ВСЁ_ДЛЯ
19. ЕСЛИ в графе $G = (V,E)$ есть циклы ТО
20. выдать "НЕТ"
21. ИНАЧЕ
22. выдать "ДА"
23. ВСЁ_ДЛЯ

Алгоритм читает потенциал слева направо по одной валентности. Если он находит левую валентность типа j, то он запоминает её позицию в j-м стеке (строка 5). Если алгоритм находит правую валентность, то он проверяет есть ли соответствующая ей левая валентность. Если она есть в позиции l, то алгоритм добавляет пару (i,l) в множество зависимостей V (строка 9). После этого он проверяет, какие из уже найденных зависимостей могут взаимодействовать с новой через функцию запретов (строки 10-14). Когда все зависимости найдены, алгоритм проверяет, есть ли в графе (V,E) циклы.

Так как длина потенциала равна n, то число зависимостей равно $n/2$. Поэтому $|V| \leqslant n/2$, $|E| \leqslant N^2/4$ и цикл в строках 10-14 работает $O(n)$. Следовательно, цикл в строках 3-18 имеет сложность $O(n^2)$. Наличие циклов в графе проверяется за время $O(\max(|V|,|E|))$. Поэтому временная сложность алгоритма составляет $O(n^2)$. □

Таким образом, справедлива следующая теорема.

Теорема 6. *Проблема принадлежности для ммКГЗ лежит в классе NP.*

Доказательство. Недетерминированный полиномиальный алгоритм для проверки принадлежности $w \in L(G)$ будет следующим.
1) Угадать строку категорий $\alpha_1^{\theta_1} \dots \alpha_n^{\theta^n} \in \delta(w)$.
2) Проверить, что $\alpha_1 \dots \alpha_n \vdash^* [S]$.
3) Проверить, что потенциал сбалансирован.□

Теорема 7. *Существует ммКГЗ G такая, что проблема принадлежности $w \in L(G)$ является NP-полной*

Доказательство. Мы докажем NP-трудность путём сведения проблемы выполнимости КНФ к проблеме принадлежности. Пусть $\Phi = C_1 \wedge \dots \wedge C_m$ — КНФ,

содержащая переменные $x_1, \ldots, x_n$. Построим ммКГЗ $G =< W, \mathbf{C}, S, \delta, \pi >$ следующим образом.
$W = \{x, \bar{x}, y, b, f, \phi\}$
Буква x соответствует вхождению в дизъюнкт некоторой переменной x_i, буква $\bar{x}$ — вхождению $\neg x_i$, а y означает, что переменная x_i не входит в дизъюнкт. Номера i этих переменных определяются положением букв x, $\bar{x}$, y в представлении дизъюнкта, поэтому индексы не нужны. Каждому дизъюнкту C_i поставим в соответствие слово $g(C_i) = z_1 \ldots z_n$, где $z_j = x$, если x_j входит в C_i, $z_j = \bar{x}$, если $\neg x_j$ входит в C_i, и $z_j = y$, если x_j не входит в C_i. Например, дизъюнкту $x_2 \vee \neg x_4 \vee x_7$ соответствует слово $yxy\bar{x}yyx$ (если КНФ содержит переменные от x_1 до x_7). В слове, которое мы построим по Φ, первый дизъюнкт будет перевёрнут. Следующий будет перевёрнут ещё раз, т. е. будет написан без изменений, третий снова будет перевёрнут и т. д. Слово будет таким: $w_\Phi = \phi b^n g(C_1)^{-1} * g(C_2) * \ldots * g(C_m)^{(-1)} * f^n$. Здесь v^{-1} — перевёрнутое v (последний дизъюнкт будет перевёрнут, если его номер нечётный).
$\mathbf{C} = \{S, 0, 1, 0', 1', A, B, T\}$
Далее определим ограничения на стрелки: $\pi(0) = \{\,1\,\}, \pi(1) = \{\,0\,\}$, $\pi(0') = \{\,1'\,\}, \pi(1') = \{\,0'\,\}$.
Теперь опишем словарь.
$\phi \mapsto [S]$
$b \mapsto [\varepsilon]^{\nearrow 0}, [\varepsilon]^{\nearrow 1}$
$x \mapsto [A\backslash T/A]^{\searrow 1 \nearrow 1'}, [A\backslash A]^{\searrow 1 \nearrow 1'}, [A\backslash A]^{\searrow 0 \nearrow 0'}, [A/A]^{\searrow 1 \nearrow 1'}, [A/A]^{\searrow 0 \nearrow 0'},$
$[A\backslash T]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 0 \nearrow 0'}, [T/A]^{\searrow 1 \nearrow 1'},$
$[B\backslash T/B]^{\searrow 1' \nearrow 1}, [B\backslash B]^{\searrow 1' \nearrow 1}, [B\backslash B]^{\searrow 0' \nearrow 0}, [B/B]^{\searrow 1' \nearrow 1}, [B/B]^{\searrow 0' \nearrow 0},$
$[B\backslash T]^{\searrow 1' \nearrow 1}, [B]^{\searrow 1' \nearrow 1}, [B]^{\searrow 0' \nearrow 0}, [T/B]^{\searrow 1' \nearrow 1}$
$\bar{x} \mapsto [A\backslash T/A]^{\searrow 0 \nearrow 0'}, [A\backslash A]^{\searrow 1 \nearrow 1'}, [A\backslash A]^{\searrow 0 \nearrow 0'}, [A/A]^{\searrow 1 \nearrow 1'}, [A/A]^{\searrow 0 \nearrow 0'},$
$[A\backslash T]^{\searrow 0 \nearrow 0'}, [A]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 0 \nearrow 0'}, [T/A]^{\searrow 0 \nearrow 0'},$
$[B\backslash T/B]^{\searrow 0' \nearrow 0}, [B\backslash B]^{\searrow 1' \nearrow 1}, [B\backslash B]^{\searrow 0' \nearrow 0}, [B/B]^{\searrow 1' \nearrow 1}, [B/B]^{\searrow 0' \nearrow 0},$
$[B\backslash T]^{\searrow 0' \nearrow 0}, [B]^{\searrow 1' \nearrow 1}, [B]^{\searrow 0' \nearrow 0}, [T/B]^{\searrow 0' \nearrow 0}$
$y \mapsto [A\backslash A]^{\searrow 1 \nearrow 1'}, [A\backslash A]^{\searrow 0 \nearrow 0'}, [A/A]^{\searrow 1 \nearrow 1'}, [A/A]^{\searrow 0 \nearrow 0'}, [A]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 0 \nearrow 0'},$
$[B\backslash B]^{\searrow 1' \nearrow 1}, [B\backslash B]^{\searrow 0' \nearrow 0}, [B/B]^{\searrow 1' \nearrow 1}, [B/B]^{\searrow 0' \nearrow 0}, [B]^{\searrow 1' \nearrow 1}, [B]^{\searrow 0' \nearrow 0}$
$* \mapsto [T\backslash \varepsilon]$
$f \mapsto [\varepsilon]^{\searrow 0}, [\varepsilon]^{\searrow 0'}, [\varepsilon]^{\searrow 1}, [\varepsilon]^{\searrow 1'}$

Утверждение. Φ выполнима тогда и только тогда, когда $w_\Phi \in L(G)$.
[$\Rightarrow$] Пусть Φ выполнима и σ — выполняющее присваивание. Присвоим буквам b категории, задающие это присваивание: i-я буква b получит категорию $[\varepsilon]^{\nearrow \sigma(x_i)}$. Каждый блок, представляющий дизъюнкт, ловит слева все стрелки, описывающие значения переменных. Например, первый блок поймает их непосредственно от b. При этом стрелки не перепутаются, так как 0 и 1 не пересекаются. Первой будет поймана стрелка, соответствующая x_n, затем стрелка, соответствующая x_{n-1}, и т. д. Поэтому первый блок записан в перевёрнутом виде. Нам нужно гарантировать, что в каждом дизъюнкте есть истинный литерал. Значения остальных литералов в этом же дизъюнкте нас не интересуют, но их нужно помнить. Допустим, что переменная x_i истинна и делает дизъюнкт истинным. Присвоим соответствующей букве x категорию $[A\backslash T/A]^{\searrow 1 \nearrow 1'}$. Теперь другие значения роли не играют, задающие их стрелки нужно поймать и послать дальше направо без изменения. Для этого в грамматике есть категории $[A\backslash A]^{\searrow 1 \nearrow 1'}, [A\backslash A]^{\searrow 0 \nearrow 0'}$ для

переменных левее x_i и $[A/A]^{\searrow 1 \nearrow 1'}, [A/A]^{\searrow 0 \nearrow 0'}$ для переменных правее. Для случая, когда x стоит на первом или на последнем месте нужны другие категории: $x \mapsto [A\backslash T]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 1 \nearrow 1'}, [A]^{\searrow 0 \nearrow 0'}, [T/A]^{\searrow 1 \nearrow 1'}$. С $\bar{x}_i$ поступим аналогично, но чтобы $\bar{x}_i$ сделал дизъюнкт истинным, x_i должен быть ложным. Значения y_i не важны, их просто пересылаем направо. Категории T нет в списках зависимостей у x, $\bar{x}$ и y, поэтому локальные категории подслова, представляющего дизъюнкт, сократятся до T. Эту категорию сократим с помощью категории $[T\backslash\varepsilon]$ символа $*$. Теперь стрелки, задающие значения переменных, идут в порядке от n к 1. Поэтому они будут пойманы в порядке от 1 к n. Построение для чётных дизъюнктов полностью аналогично построению для нечётных дизъюнктов. Используются категории, содержащие B вместо A. Они ловят $0'$ и $1'$, а выпускают 0 и 1, а переменные записаны от 1 до n. Блок из символов f в конце слова просто поймает все стрелки. Наконец, присвоим категорию $[S]$ символу ϕ. Таким образом, мы сопоставили слову w_Φ строку категорий Γ такую, что $\Gamma \vdash^* [S]$. Это значит, что $w_\Phi \in L(G)$.
$[\Leftarrow]$ Пусть Φ не выполнима и пусть b задают какое-нибудь присваивание. Пусть C_j — первый ложный дизъюнкт. До него все значения переменных доберутся, как и выше. Локальные категории дизъюнкта должны сократиться до ε. Это значит, что некоторый x или $\bar{x}$ должен получить категорию с T. Но такая категория для x_i ловит слева 1, а $x_i = 0$ (иначе дизъюнкт C_j был бы истинным). Аналогично категория для $\bar{x}_i$ ловит слева 0, а $x_i = 1$. В обоих случаях стрелки у x не сократятся. Сократиться же с другими стрелками они не могут, так как 0 и 1 не пересекаются. Получилось, что $w \notin L(G)$. $\Box$

Грамматика G, построенная в доказательстве теоремы, содержит четыре типа дальних связей. С помощью некоторых технических усложнений этот результат можно усилить.

Теорема 8. *Существует ммКГЗ G с двумя типами валентностей такая, что проблема принадлежности для G является NP-полной.*

Набросок доказательства.
Как и в предыдущей теореме, NP-трудность доказывается путём сведения проблемы выполнимости к проблеме принадлежности. Символы x, $\bar{x}$ и y ловили дальнюю зависимость слева и выпускали новую дальнюю зависимость направо. При этом не использовались локальные части категорий. Когда символ x, $\bar{x}$ или y из первого блока ловит стрелку слева в новой грамматике, значение этой стрелки будет запоминаться в локальной части категории. Категории второго блока проверяют, что значения переменных сохраняются, и выпускают направо соответствующие стрелки. Грамматика будет следующей.
$W = \{\, b, f, \phi, x, \bar{x}, y, *, \# \,\}$
$\mathbf{C} = \{\, S, 0, 1, T, L, R, E, X_0, X_1, X_0', X_1', Y_0, Y_1 \,\}$
Словарь δ определим следующим образом (категории для x, $\bar{x}$ и y получены с помощью Γ).
$b \mapsto [\varepsilon]^{\nearrow 0}, [\varepsilon]^{\nearrow 1}$
$f \mapsto [\varepsilon]^{\searrow 0}, [\varepsilon]^{\searrow 1}$
$\phi \mapsto [S]$
$x \mapsto [T/X_0/T]^{\searrow 0}, [T/X_0'/R]^{\searrow 0}, [T/Y_0/T]^{\searrow 0}, [T/X_1/E]^{\searrow 1}, [T/X_1'/L]^{\searrow 1}, [T/Y_1/L]^{\searrow 1},$
$[L/X_0/L]^{\searrow 0}, [L/X_0'/E]^{\searrow 0}, [L/Y_0/L]^{\searrow 0}, [L/X_1/E]^{\searrow 1}, [L/X_1'/L]^{\searrow 1}, [L/Y_1/L]^{\searrow 1},$
$[R/X_0/R]^{\searrow 0}, [R/X_0'/R]^{\searrow 0}, [R/Y_0/R]^{\searrow 0}, [R/X_1/E]^{\searrow 1}, [R/X_1'/E]^{\searrow 1}, [R/Y_1/E]^{\searrow 1},$

$[E/X_0/E]^{\searrow 0}, [E/X_0'/E]^{\searrow 0}, [E/Y_0/E]^{\searrow 0}, [E/X_1/E]^{\searrow 1}, [E/X_1'/E]^{\searrow 1}, [E/Y_1/E]^{\searrow 1},$
$[X_0]^{\nearrow 0}, [X_1]^{\nearrow 1}$
$\bar{x} \mapsto [T/X_0/L]^{\searrow 0}, [T/X_0'/E]^{\searrow 0}, [T/Y_0/L]^{\searrow 0}, [T/X_1/R]^{\searrow 1}, [T/X_1'/T]^{\searrow 1}, [T/Y_1/T]^{\searrow 1},$
$[L/X_0/L]^{\searrow 0}, [L/X_0'/E]^{\searrow 0}, [L/Y_0/L]^{\searrow 0}, [L/X_1/E]^{\searrow 1}, [L/X_1'/L]^{\searrow 1}, [L/Y_1/L]^{\searrow 1},$
$[R/X_0/E]^{\searrow 0}, [R/X_0'/E]^{\searrow 0}, [R/Y_0/E]^{\searrow 0}, [R/X_1/R]^{\searrow 1}, [R/X_1'/R]^{\searrow 1}, [R/Y_1/R]^{\searrow 1},$
$[E/X_0/E]^{\searrow 0}, [E/X_0'/E]^{\searrow 0}, [E/Y_0/E]^{\searrow 0}, [E/X_1/E]^{\searrow 1}, [E/X_1'/E]^{\searrow 1}, [E/Y_1/E]^{\searrow 1},$
$[X_0']^{\nearrow 0}, [X_1']^{\nearrow 1}$
$y \mapsto [T/X_0/T]^{\searrow 0}, [T/X_0'/R]^{\searrow 0}, [T/Y_0/T]^{\searrow 0}, [T/X_1/R]^{\searrow 1}, [T/X_1'/R]^{\searrow 1}, [T/Y_1/R]^{\searrow 1},$
$[L/X_0/L]^{\searrow 0}, [L/X_0'/E]^{\searrow 0}, [L/Y_0/L]^{\searrow 0}, [L/X_1/E]^{\searrow 1}, [L/X_1'/L]^{\searrow 1}, [L/Y_1/L]^{\searrow 1},$
$[R/X_0/R]^{\searrow 0}, [R/X_0'/R]^{\searrow 0}, [R/Y_0/R]^{\searrow 0}, [R/X_1/R]^{\searrow 1}, [R/X_1'/R]^{\searrow 1}, [R/Y_1/R]^{\searrow 1},$
$[E/X_0/E]^{\searrow 0}, [E/X_0'/E]^{\searrow 0}, [E/Y_0/E]^{\searrow 0}, [E/X_1/E]^{\searrow 1}, [E/X_1'/E]^{\searrow 1}, [E/Y_1/E]^{\searrow 1},$
$[Y_0]^{\nearrow 0}, [Y_1]^{\nearrow 1}$
$* \mapsto [E]$
$\# \mapsto [\varepsilon/T]$
Теперь определим ограничения π.
$\pi(0) = \{ 1 \}, \pi(1) = \{ 0 \}$

Пусть $\Phi = C_1 \wedge \dots \wedge C_{2m}$ — КНФ. Можно считать, что она содержит чётное число дизъюнктов. Слово, кодирующее КНФ, будет следующим:
$w_\Phi = \phi b^n \# g(C_1)^{-1} * g(C_2) \# g(C_3)^{-1} * g(C_4) \dots \# g(C_{2m-1})^{-1} * g(C_{2m}) f^n$. $\square$

Пусть, например, задана КНФ $\Phi = (x_1 \vee x_2) \wedge (\neg x_2 \vee \neg x_3)$, а переменные имеют значения $\sigma(x_1) = 1$, $\sigma(x_2) = 1$, $\sigma(x_3) = 0$. Эта КНФ представляется словом $\phi bbb \# yxx * y\bar{x}\bar{x}fff$. Этому слову можно приписать следующую строку категорий: $[S][\varepsilon]^{\nearrow 1}[\varepsilon]^{\nearrow 1}[\varepsilon]^{\nearrow 0}[\varepsilon/T][T/X_0'/R]^{\searrow 0}$ $[R/X_1'/E]^{\searrow 1}[E/Y_1'/E]^{\searrow 1}[E][Y_1]^{\nearrow 1}[X_1']^{\nearrow 1}[X_0']^{\nearrow 0}$ $[\varepsilon]^{\searrow 0}[\varepsilon]^{\searrow 1}[\varepsilon]^{\searrow 1}$.

Предположим теперь, что в ммКГЗ $G = < W, \mathbf{C}, S, \delta, \pi >$ только один тип непроективных зависимостей участвует в ограничениях π, например, X. Имеется три возможных варианта.

1) $\pi(A) = \varnothing$ для всех $A \neq X$ (т.е. только $\pi(X)$ может быть непустым). Пусть θ — потенциал такой, что $\theta \vdash^* \varepsilon$, если не использовать ограничения. Этот потенциал останется сбалансированным, если использовать π. Сначала мы можем сократить валентности, тип которых A отличен от X. Это возможно потому, что $\pi(A) = \varnothing$. Оставшийся потенциал будет содержать только X, так что π не сможет запретить какие-либо сокращения.

2) $\pi(A) \subseteq \{ X \}$ для всех A (т.е. единственный тип, который может предотвратить некоторые сокращения, — это X). Этот случай аналогичен предыдущему. Сначала мы можем сократить все валентности X. Если $\pi(X) = \{ X \}$, это не повлияет на возможность таких сокращений, так как это ограничение уже определено правилом дальней зависимости. Потом мы сможем сократить все остальные валентности. Таким образом, грамматики, удовлетворяющие условиям одного из двух первых случаев, эквивалентны оКГЗ без ограничений π.

3) $\pi(A) \subseteq \{ X \}$ для всех $X \neq A$ (т.е. X может запретить некоторые сокращения, и некоторые валентности могут запретить сокращение X). Грамматика, построенная в доказательстве предыдущей теоремы, удовлетворяет этому условию. Поэтому даже если всего один тип дальних зависимостей участвует в ограничениях, проблема принадлежности остаётся NP-полной.

Благодарности. Автор благодарен М.И. Дехтярю за постановки задач и внимание к работе и А.Я. Диковскому за полезное обсуждение результатов.

Заключение

В этой статье мы исследовали расширение категориальных грамматик зависимостей — мультимодальные категориальные грамматики зависимостей (ммКГЗ) — и изучили свойства класса распознаваемых ими языков. Раздел 2 был посвящён свойствам замкнутости класса ммКГЗ-языков. В теореме 1 была установлена замкнутость класса ммКГЗ-языков относительно итерации. Результаты о замкнутости класса оКГЗ-языков относительно операций объединения, конкатенации, пересечения с регулярными языками, неукорачивающего гомоморфизма и обращения гомоморфизма непосредственно переносятся на класс ммКГЗ-языков. На основании этих результатов в теореме 2 сделан вывод, что класс ммКГЗ-языков образует абстрактное семейство языков. Также мы установили, что класс ммКГЗ-языков замкнут относительно пересечения, но не замкнут относительно проекции. В разделе 4 мы привели пример неполулинейного языка, задаваемого ммКГЗ. В разделе 5 рассмотрена проблема принадлежности для ммКГЗ. Мы показали, что для ммКГЗ эта проблема является NP-полной для одной грамматики и распознаваемого ей языка.

Список литературы

[1] Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1. М.: Мир, 1978.

[2] Гинзбург С. Математическая теория контекстно-свободных языков. М.: Мир, 1970.

[3] Гладкий А.В. Формальные грамматики и языки. М.: Наука, 1973.

[4] Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. М.: Наука, 1969.

[5] Карлов Б.Н. Нормальные формы и автоматы для категориальных грамматик зависимостей // Вестник Тверского государственного университета, серия «Прикладная математика», 2008, №35 (95). С. 23-43.

[6] Bar-Hillel Y., Gaifman H., Shamir E. On categorial and phrase structure grammars. Bull. Res. Council Israel, 9F, 1960, pp. 1–16.

[7] Dekhtyar M., Dikovsky A. Categorial Dependency Grammars // Proc. of Int. Conf. on Categorial Grammars, 2004.

[8] Dekhtyar M., Dikovsky A. Generalized Categorial Dependency Grammars // Pillars of Computer Science: Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday. LNCS, v. 4800, 2008.

[9] Dikovsky A. Architecture compositionnelle pour les dépendances croisées // TALN 2007, Toulouse, 2007.

[10] Ginsburg S., Greibach S.A., Harrison M.A. One-way stack automata. Journal of The ACM - JACM , vol. 14, no. 2, pp. 389-418, 1967.

[11] Ginsburg S., Greibach S. Abstract families of languages. Mem. Amer. Math. Soc., 87 (1969), pp. 1-32.

[12] Tesnière L. Éléments de syntaxe structurale. Librairie C. Klincksieck, Paris, 1959.